

Creating National Air Pollution Models for Population Exposure Assessment in Canada

Perry Hystad,¹ Eleanor Setton,² Alejandro Cervantes,³ Karla Poplawski,⁴ Steeve Deschenes,² Michael Brauer,⁴ Aaron van Donkelaar,⁵ Lok Lamsal,⁵ Randall Martin,^{5,6} Michael Jerrett,⁷ and Paul Demers^{4,8}

¹School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada; ²Department of Geography, University of Victoria, Victoria, British Columbia, Canada; ³Department of Geography, and ⁴School of Environmental Health, University of British Columbia, Vancouver, British Columbia, Canada; ⁵Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada; ⁶Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA; ⁷School of Public Health, Division of Environmental Health Science, University of California–Berkeley, Berkeley, California, USA; ⁸Occupational Cancer Research Centre, Cancer Care Ontario, Toronto, Ontario, Canada

BACKGROUND: Population exposure assessment methods that capture local-scale pollutant variability are needed for large-scale epidemiological studies and surveillance, policy, and regulatory purposes. Currently, such exposure methods are limited.

METHODS: We created 2006 national pollutant models for fine particulate matter [PM with aerodynamic diameter $\leq 2.5 \mu\text{m}$ ($\text{PM}_{2.5}$)], nitrogen dioxide (NO_2), benzene, ethylbenzene, and 1,3-butadiene from routinely collected fixed-site monitoring data in Canada. In multiple regression models, we incorporated satellite estimates and geographic predictor variables to capture background and regional pollutant variation and used deterministic gradients to capture local-scale variation. The national NO_2 and benzene models are evaluated with independent measurements from previous land use regression models that were conducted in seven Canadian cities. National models are applied to census block-face points, each of which represents the location of approximately 89 individuals, to produce estimates of population exposure.

RESULTS: The national NO_2 model explained 73% of the variability in fixed-site monitor concentrations, $\text{PM}_{2.5}$ 46%, benzene 62%, ethylbenzene 67%, and 1,3-butadiene 68%. The NO_2 model predicted, on average, 43% of the within-city variability in the independent NO_2 data compared with 18% when using inverse distance weighting of fixed-site monitoring data. Benzene models performed poorly in predicting within-city benzene variability. Based on our national models, we estimated Canadian ambient annual average population-weighted exposures (in micrograms per cubic meter) of 8.39 for $\text{PM}_{2.5}$, 23.37 for NO_2 , 1.04 for benzene, 0.63 for ethylbenzene, and 0.09 for 1,3-butadiene.

CONCLUSIONS: The national pollutant models created here improve exposure assessment compared with traditional monitor-based approaches by capturing both regional and local-scale pollution variation. Applying national models to routinely collected population location data can extend land use modeling techniques to population exposure assessment and to informing surveillance, policy, and regulation.

KEY WORDS: air pollution, Canada, fixed-site monitors, gradients, land use regression, population exposure assessment, satellite data. *Environ Health Perspect* 119:1123–1129 (2011). doi:10.1289/ehp.1002976 [Online 31 March 2011]

Predicting air pollution concentrations at resolutions capable of capturing local-scale pollutant gradients over large geographical areas is becoming increasingly important in multicity and national health studies; in population exposure assessment; and in support of policy, surveillance, and regulatory initiatives. Currently, fixed-site government monitors are the foundation of these activities; however, because of siting criteria, such monitors may fail to fully capture local-scale pollutant variability. In addition, the number of monitors and their spatial distribution may be limited, as is the case in Canada. At present, few methodologies are available that adequately capture local-scale pollutant variability at a national scale when monitor density, distribution, or siting is suboptimal.

A number of approaches may be used to model air pollution over large areas, including interpolation of fixed-site government monitoring data, dispersion modeling, satellite remote sensing, land use regression (LUR),

and proximity and deterministic methods. Each approach, however, has inherent limitations that restrict its use for producing local-scale pollution estimates. Interpolation of fixed-site air pollution monitoring data has typically been used to predict pollution concentrations across large areas (Beelen et al. 2009), with recent interest directed towards kriging methods and spatial smoothing with geographic covariates (Beelen et al. 2009; Hart et al. 2009; Yanosky et al. 2008). Fixed-site monitors may not capture entire populations, and measurements typically represent regional and between-city pollution differences due to monitor siting criteria, which prevent monitors from being placed in proximity to major roads and other pollution sources. Dispersion models also exist for large geographical areas and have been incorporated into regulatory and epidemiological studies of air pollution (Cyrus et al. 2005; Nafstad et al. 2003). Importantly, the resolutions of pollutant estimates from

dispersion models over large geographical areas are typically restricted, for example, to 1 or 3 km^2 (Jerrett et al. 2005). Satellite remote sensing is a new methodology available to predict air pollution concentrations over large geographical areas, and a number of studies have evaluated different remotely sensed concentrations of fine particulate matter [PM with aerodynamic diameter $\leq 2.5 \mu\text{m}$ ($\text{PM}_{2.5}$)] (e.g., van Donkelaar et al. 2010) and gaseous pollutants (Martin 2008) and found moderate to good associations with ground-level monitoring data. Currently, the resolution of satellite data limits their use to representing regional pollution concentrations, but indicators of local air pollution may be used in concert to improve the spatial resolution of predictions (Liu et al. 2009). LUR approaches have been used extensively to predict within-city pollutant concentrations of nitrogen dioxide (NO_2) and $\text{PM}_{2.5}$ (for review, see Hoek et al. 2008), but to a lesser extent for volatile organic compounds (VOCs). However, the approach is well suited to modeling pollutants that exhibit significant spatial variation, especially traffic-related VOCs (Atari and Luginaah 2009; Mukerjee et al. 2009; Smith et al. 2006; Su et al. 2010; Wheeler et al. 2008). The city-by-city approach in which LUR models are created is costly, and integration and interpretation across multiple city models is difficult. Simple proximity and deterministic approaches have also been widely used as surrogates for exposure to vehicle and industrial sources,

Address correspondence to P. Hystad, School of Population and Public Health, University of British Columbia, 2206 East Mall, Vancouver, BC, V6T 1Z3 Canada. Telephone: (604) 312-4768. Fax: (604) 822-9588. E-mail: phystad@gmail.com

Supplemental Material is available online (doi:10.1289/ehp.1002976 via <http://dx.doi.org/>).

We thank R. Allen for providing the Edmonton and Winnipeg monitoring data; D. Crouse, M. Goldberg, and N. Ross for the Montreal data; and O. Atari for the Sarnia data.

The research was supported by a grant from the Canadian Partnership against Cancer. Health Canada also provided support for the development of the satellite-derived pollution estimates.

The authors declare they have no actual or potential competing financial interests.

Received 15 September 2010; accepted 31 March 2011.

specifically in epidemiological studies; yet, such measures in isolation are often poor surrogates for exposure. To date, few population exposure assessments have incorporated multiple sources of data, specifically satellite pollutant estimates, LUR modeling of geographic characteristics, and information on proximity and pollution gradients, to estimate local-scale air pollution concentrations at a national scale.

Here we report a modeling initiative to produce 2006 national $PM_{2.5}$, NO_2 , benzene, ethylbenzene, and 1,3-butadiene models for Canada that capture local-scale pollutant variability and apply these models to routinely collected population location data to calculate population exposures. This research is part of Carex Canada, a national surveillance initiative designed to estimate the number of Canadians potentially exposed to known or suspected environmental and occupational carcinogens (Carex Canada 2011). This research adds to the literature on air pollution modeling and exposure assessment by creating national LUR models from fixed-site monitoring data; incorporating various predictor data sets and methods to capture the different scales of pollution sources;

and extending LUR modeling techniques to population exposure assessment and to inform surveillance, policy, and regulation.

Materials and Methods

Pollutant modeling approach. Models were developed in two stages using different predictor variables and methodology to capture background, regional, and local-scale pollution variation. First, for each National Air Pollution Surveillance (NAPS) fixed-site monitoring station, we derived satellite-based estimates ($PM_{2.5}$ and NO_2 only) and geographic variables (e.g., road length, population density, proximity to large emitters) using ArcGIS (version 9.3; ESRI, Redland, CA, USA). We used forward stepwise regression to develop LUR models and retained variables that corresponded to hypothesized effect directions; we maximized the sums of squares explained by Akaike's information criterion. Spatial autocorrelation was also evaluated using the Moran's I statistic in ArcGIS. We sought to develop parsimonious models rather than traditional predictive models that maximize prediction but make interpretation of individual variable

contributions difficult. Only variables significant at the $p < 0.05$ level were included in the final models. As expected, NAPS monitoring locations in Canada did not display sufficient variability to estimate model coefficients for important local-scale parameters, such as proximity to major roadways, because of monitoring siting. Thus, local-scale predictors were underpowered in the LUR modeling approach.

In the second stage, we conducted comprehensive literature reviews to identify deterministic factors to represent local-scale gradients in pollutant concentrations associated with specific sources (i.e., highways, major roads, gas stations). For each pollutant, we identified concentrations near these selected sources in relation to local background levels and developed deterministic multipliers with distance decay rates (together referred to as gradients in this paper) to apply to the background and regional concentrations predicted by our LUR models. All statistical analyses were conducted using SAS (version 9.1; SAS Institute Inc., Cary, NC, USA).

Air quality data. Annual average concentrations of $PM_{2.5}$ (177 monitoring stations), NO_2 (134 monitors), and benzene, ethylbenzene, and 1,3-butadiene (53 monitors) were calculated using data from unique NAPS monitoring sites that were operating during 2006 (see Figure 1). Continuous monitoring data from a given monitor were included if at least 50% of hourly observations were available for a 24-hr period and at least 50% of days were available in a month. Monthly averages from filter-based $PM_{2.5}$ measurements required a minimum of three of five valid measurements per month. Annual averages for 2006 were not calculated for individual monitors unless there were at least 6 months of complete data with one valid month per quarter.

NAPS includes different monitor types for $PM_{2.5}$, including tapered element oscillating microbalances (TEOMs), dichotomous particulate samplers (Thermo Fisher Scientific Inc., Waltham, MA, USA), and beta-attenuation mass monitors (Met One Instruments Inc., Grants Pass, OR, USA). Multiple monitors are often present at one location, and our comparative analysis found differences in levels measured by TEOMs, which are known to underpredict $PM_{2.5}$ because of nitrate evaporation (Dann T, personal communication). We therefore selected other monitor types when they were available at the same location. Those stations with only TEOMs available were adjusted based on yearly calibration between collocated dichotomous and TEOM monitors during 2006 [$n = 14$, dichotomous = $1.640 + 1.089 \times (\text{TEOM})$, $R^2 = 0.89$, $p < 0.001$]. NO_2 , benzene, ethylbenzene, and 1,3-butadiene were measured using standard methods (NAPS 2004).

Predictor variables. $PM_{2.5}$ and NO_2 satellite data. Canada-wide concentrations of

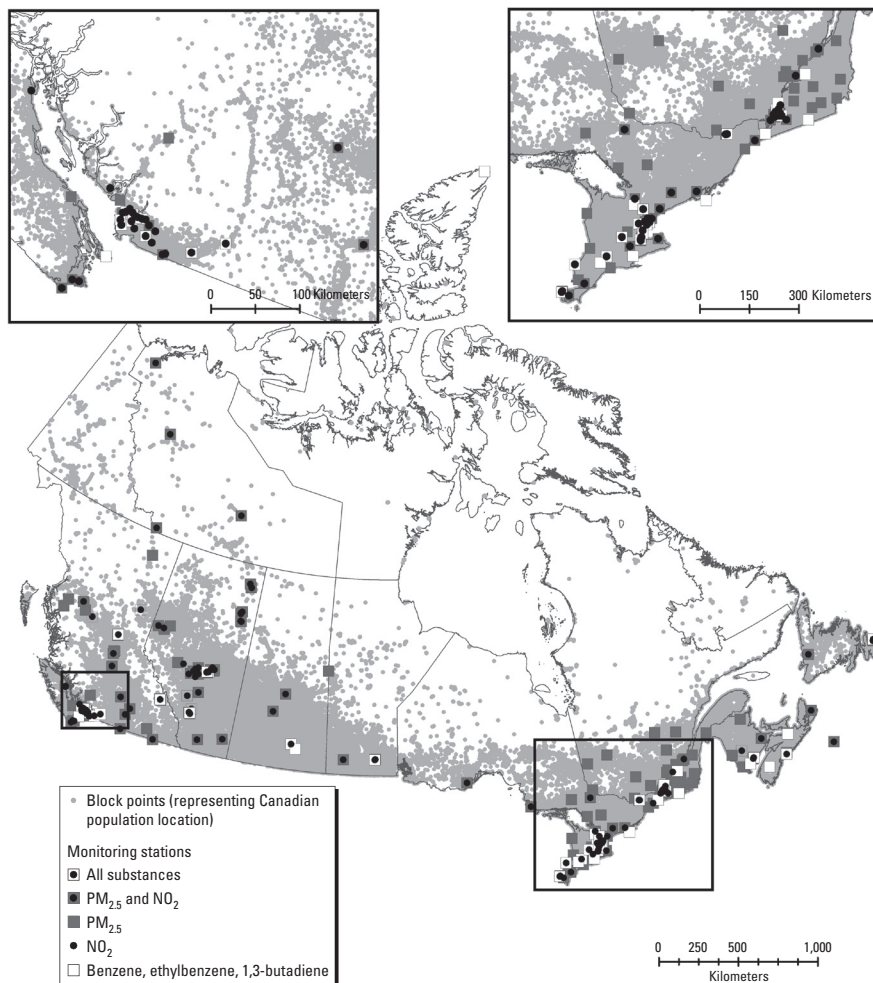


Figure 1. Location of NAPS monitors that were used to create national $PM_{2.5}$, NO_2 , benzene, ethylbenzene, and 1,3-butadiene models.

PM_{2.5} and NO₂ were estimated using satellite atmospheric composition data combined with local, coincident scaling factors from a chemical transport model [Goddard Earth Observing System (GEOS)-Chem 2011]. Ground-level PM_{2.5} estimates were derived from aerosol optical depth data from the Terra satellite [National Aeronautics and Space Administration (NASA) 2011b], in combination with output from GEOS-Chem simulations to estimate the relationship between aerosol optical depth over the atmospheric column and ground-level PM_{2.5} (van Donkelaar et al. 2010). Ground-level NO₂ concentrations were estimated from tropospheric NO₂ columns retrieved from the ozone monitoring instrument on the Aura satellite (NASA 2011a); GEOS-Chem was also used to calculate the relationship between the NO₂ column and ground-level concentration (Lamsal et al. 2008). Both PM_{2.5} and NO₂ were estimated at a 0.1 × 0.1° resolution (~ 10 × 10 km). Estimates for PM_{2.5} were calculated from 2001–2006 data to ensure sufficient observations. For NO₂ estimates, we used data from 2005 and 2006, because ozone monitoring instrument measurements began in late 2004.

Geographic data. We modeled regional pollutant variation using geographic predictor variables potentially relevant to pollutant sources, emissions, and dispersion. To capture varying spatial influences of predictors, all variables were calculated for circular buffer distances ranging from 50 m to 50 km. Classes of variables included population density derived from census block-face points (Statistics Canada 2006); 1-km land use classifications (Global Land Cover Characterization 2008); high-resolution (30 m) land-use classifications (DMTI Spatial Inc., Markham, Ontario, Canada); sources of large industrial emissions from the Canadian National Pollutant Release Inventory (NPRI; Environment Canada 2010); small point source locations extracted from the Dun and Bradstreet (D&B) Selectory database of businesses (Hoovers, Austin, TX, USA) in Canada; length of and distance to specific road classifications using the DMTI Spatial road network, such as freeway, highway, major road, and minor road (DMTI Spatial Inc.); length and density of railroads; elevation; and meteorological variables (precipitation and temperature). Any geographic variables with > 30% zero values—those with no predictive features in proximity to a monitor—were recoded as binary (i.e., present/absent). In total, 10 variable classes and 270 buffer-specific variables were explored in the LUR models.

Deterministic gradients. Gradients were developed with a focus on mobile sources and gas stations. We conducted a comprehensive literature review of published studies to identify the distance from sources at which

pollutant concentrations typically return to background levels, and an expected ratio of near-source pollutant levels compared with background pollutant levels for each source and pollutant. We searched PubMed (2010), Web of Science (Thomson Reuters 2010), and Google Scholar (2010) using a range of keywords to identify studies with measurements of pollutant gradients. Studies varied widely in terms of location, date, methods, duration of measures, number of samples, and definition of near source and background. We developed linear gradients using the steepest portion of the exponential decay curves typically found in the literature, as the tails of the decay functions were very sensitive to local parameters. Gradients were also selected to represent Canadian conditions. Table 1 summarizes the gradients developed for Canada and applied to the LUR models.

To identify the distance of each NAPS monitor from the nearest highway, major road, local urban road, and gas station, we used DMTI road network data and D&B commercial data for point sources. If a monitor was close enough to one of these features for the source to influence pollutant levels, we modified the corresponding LUR model results (not including point source industrial variables) to account for the deterministic gradients. For example, based on our review of the literature, we assumed that NO₂ concentrations at the side of a highway would be 1.65 times higher than LUR-based background concentrations but consistent with background levels 300 m from the highway; this assumption resulted in a distance decay rate of 0.33% per meter that was applied to the model to estimate NO₂ levels within the 300-m gradient buffer.

Model evaluation. We used three approaches for model evaluation. Due to the small number of NAPS monitoring stations for PM_{2.5}, NO₂, benzene, ethylbenzene, and 1,3-butadiene, we did not leave out a percentage for independent postmodel evaluation, because we wanted to capture the greatest range of

model predictors possible. Therefore, we first evaluated all LUR models using a bootstrap approach to determine the sensitivity of model prediction and parameter estimates to monitor sampling. Random selection of monitors was conducted, with replacement, and variable coefficients and model *R*² values were recorded from the new full sample. This was repeated for 10,000 iterations to estimate the 95% confidence interval (CI) for overall model prediction and individual variable coefficients. Next, we conducted a leave-one-out analysis where each LUR model was repeatedly parameterized on *n* – 1 data points and then used to predict the excluded monitor measurement. The mean differences between the predicted and measured values were used to estimate model error.

Finally, we evaluated the NO₂ and benzene LUR models, with and without gradients, against independent data (35–196 monitoring sites per city) previously collected for LUR models in seven Canadian cities (for a full description of data collection and modeling see Allen et al. 2010; Atari and Luginaah 2009; Crouse et al. 2009; Henderson et al. 2007; Jerrett et al. 2007; Su et al. 2010). Briefly, in each city, monitoring took place over a 2-week period; data from fixed-site monitors, monitoring during yearly average concentration periods, or multiple measurement periods were used to estimate yearly averages [see Supplemental Material, Table 1 (doi:10.1289/ehp.1002976) for the city-specific data used for model evaluation]. These pollution measurements were collected at much higher spatial densities than were NAPS and from monitors that were located to specifically capture spatial pollutant gradients. Consequently, these data were reasonable for use as a gold standard to determine how well the two national NO₂ and benzene models (the LUR models and the LUR models with gradients) predicted within-city variation. In addition, we compared the city-specific data with estimates based on inverse distance weighting (IDW) of annual average NO₂ and benzene concentrations measured at NAPS monitors

Table 1. PM_{2.5}, NO₂, benzene, ethylbenzene, and 1,3-butadiene gradients determined from the literature and incorporated with national LUR model predictions.

| Substance | Source | Increase at source | Gradient distance (m) |
|-------------------|--------------------|--------------------|-----------------------|
| PM _{2.5} | Highway | 1.25 ^a | 75 ^b |
| | Major road | 1.1 ^a | 75 ^b |
| NO ₂ | Highway | 1.65 ^a | 300 ^c |
| | Major road | 1.2 ^a | 100 ^c |
| Benzene | Gas station | 6.5 ^d | 100 ^d |
| | Highway/major road | 3.25 ^e | 50 ^f |
| | Local road | 1.5 ^e | 50 ^f |
| Ethylbenzene | Highway | 3.7 ^g | 300 ^h |
| | Major road | 2.2 ^g | 300 ^h |
| | Local road | 1.4 ^g | 300 ^h |
| 1,3-Butadiene | Highway | 4 ⁱ | 75 ^j |

^aSmargiassi et al. (2005). ^bBeckerman et al. (2008). ^cHitchins et al. (2000). ^dRoorda-Knape et al. (1998). ^eTiitta et al. (2002). ^fBeckerman et al. (2008). ^gGilbert et al. (2003, 2007). ^hRoorda-Knape et al. (1998). ⁱSu et al. (2009). ^jKarakitsios et al. (2007). ^kHellén et al. (2006). ^lParra et al. (2009). ^mThorsson and Eliasson (2006). ⁿVardoulakis et al. (2002). ^oBeckerman et al. (2008). ^pThorsson and Eliasson (2006). ^qVenkatram et al. (2009). ^rParra et al. (2009). ^sRoukos et al. (2009). ^tWang and Zhao (2008). ^uWang and Zhao (2008).

(with and without deterministic gradients). Because of NAPS monitor density in Canada, kriging could not be applied.

Population exposure assessment. The national pollutant models were applied to each of the 478,831 Statistics Canada street block-face centroid locations in 2006 to estimate population exposures. First, we applied the LUR models to each block point to derive

a unique predicted pollutant concentration for each point, representing the average exposure level for 89 and a SD of ± 158 individuals. We used a GIS to identify the distance of each block centroid to the nearest highway, major road, local urban road, and gas stations and adjusted the corresponding LUR model estimate when the street block point was located within an associated gradient. We

then estimated population-weighted exposures to PM_{2.5}, NO₂, benzene, ethylbenzene, and 1,3-butadiene in the Canadian population as a whole, and we estimated uncertainty using the 95% confidence limits for LUR model predictions. Because there was insufficient information in the literature to examine uncertainty for specific gradients, we selected $\pm 50\%$ for all gradients (values shown in Table 1).

Results

National LUR model results. Table 2 summarizes the national LUR model results. The PM_{2.5} model predicted 46% of PM_{2.5} variation and was dominated by satellite predictions, which alone explained 41% of PM_{2.5} variation. The NO₂ model predicted 73% of NO₂ variation and length of all roads within 10 km was the dominant predictor, explaining 55% of NO₂ variation. This variable was only moderately correlated ($r = 0.56$) to NO₂ predictions from satellite data, which further explained 4% of NO₂ variation in the final model. The models for benzene, ethylbenzene, and 1,3-butadiene had similar predictive results, explaining 62, 67, and 68% of pollutant variability, respectively. Data from one monitor were removed as an outlier from the benzene and ethylbenzene models (St. John Baptiste, located in Montreal east city) and from the 1,3-butadiene model (Sarnia, located in southern Ontario near the Detroit–Windsor border), which were associated with the highest pollutant concentration for each substance.

Spatial autocorrelation of national LUR models. Spatial autocorrelation of the LUR model residuals was examined using Moran's I in ArcGIS. Spatial autocorrelation was present in the PM_{2.5} LUR model residuals (Moran's $I = 0.33$, $p < 0.001$), indicating that a moderate amount of spatial autocorrelation

Table 2. National LUR model results for PM_{2.5}, NO₂, benzene, ethylbenzene, and 1,3-butadiene.

| Variable | Distance ^a | Value | SE | p-Value |
|---|-----------------------|---------|----------|----------|
| PM _{2.5} model ($R^2 = 0.46$, RMSE = 1.529) | | | | |
| Intercept | — | 2.802 | 0.497 | < 0.0001 |
| Satellite PM _{2.5} ($\mu\text{g}/\text{m}^3$) | — | 2.392 | 0.263 | < 0.0001 |
| NPRI emissions (tonnes) | 5 km | 1.63e-3 | 5.95e-4 | 0.007 |
| Industrial land use (m^2) | 1 km | 1.03e-6 | 4.18e-7 | 0.014 |
| NO ₂ model ($R^2 = 0.73$, RMSE = 5.470) | | | | |
| Intercept | — | 13.179 | 1.374 | < 0.0001 |
| Satellite NO ₂ (ppb) | — | 1.4903 | 0.355 | < 0.0001 |
| Industrial land use (m^2) | 2 km | 3.21e-6 | 5.73e-7 | < 0.0001 |
| Road length (m) | 10 km | 7.42e-6 | 9.04e-7 | < 0.0001 |
| Summer rainfall (mm) | — | -0.010 | 0.002 | < 0.0001 |
| Benzene model ^b ($R^2 = 0.62$, RMSE = 0.298) | | | | |
| Intercept | — | 0.346 | 0.069 | < 0.001 |
| Major road length (m) | 10 km | 1.18e-6 | 2.56e-7 | < 0.001 |
| NPRI emissions (present) | 10 km | 0.526 | 0.089 | < 0.001 |
| Ethylbenzene model ^c ($R^2 = 0.67$, RMSE = 0.193) | | | | |
| Intercept | — | 0.152 | 0.039 | < 0.001 |
| Population (count) | 10 km | 6.74e-7 | 7.25e-8 | < 0.001 |
| NPRI emissions (present) | 2 km | 0.272 | 0.071 | < 0.001 |
| 1,3-Butadiene model ^d ($R^2 = 0.68$, RMSE = 0.034) | | | | |
| Intercept | — | 0.011 | 0.009 | 0.208 |
| Road length (m) | 750 m | 3.89e-6 | 7.93e-7 | < 0.001 |
| Highway (present) | 500 m | 0.041 | 0.012 | 0.002 |
| Commercial land use (m^2) | 10 km | 1.60e-9 | 5.97e-10 | 0.010 |

Satellite PM_{2.5} and NO₂ are satellite-derived estimates of PM_{2.5} and NO₂. Land use is the area of specific land-use types (industrial, commercial) within the associated buffer distance. Road length refers to the length of different road classifications (all, major, highways) within the associated buffer distance. Summer rainfall refers to the amount of rainfall recorded from May to September from the nearest meteorological station. NPRI emissions refer to the amount of annual emissions of the model substance released from industries that reported to the NPRI. NPRI emissions (present) refers to the presence of NPRI facilities that have released a model substance into the air. Population (count) refers to the number of individuals who resided within the associated buffer distance.

^aRadius of circular buffers used to derive variables. ^bOne outlier removed with benzene concentration of $3.55 \mu\text{g}/\text{m}^3$. ^cOne outlier removed with ethylbenzene concentration of $2.57 \mu\text{g}/\text{m}^3$. ^dOne outlier removed with 1,3-butadiene concentration of $0.82 \mu\text{g}/\text{m}^3$.

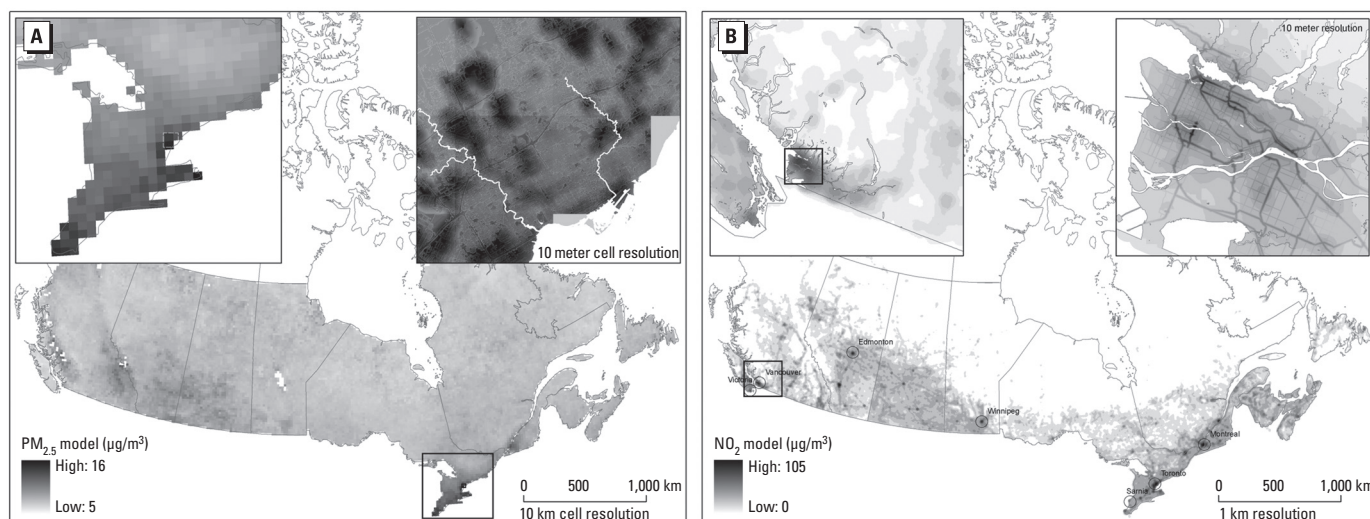


Figure 2. National annual average models for PM_{2.5}, highlighting southern Ontario and the city of Toronto (A), and for NO₂, highlighting southwestern British Columbia and the city of Vancouver (B), that incorporate satellite-derived pollutant estimates, geographic land use variables, and deterministic gradients. The seven cities shown in (B) represent locations of independent monitoring data used to evaluate the national NO₂ and benzene models.

remained that was not explained by the $PM_{2.5}$ model predictors. Clustering of positive residuals (model underpredicting by an average of $2.57 \mu\text{g}/\text{m}^3$) occurred in the rural interior of British Columbia. An indicator variable for British Columbia substantially reduced the spatial autocorrelation (Moran's $I = 0.03$, $p = 0.04$). Sensitivity analysis using a summer-only $PM_{2.5}$ model indicated no spatial autocorrelation (Moran's $I = 0.04$, $p = 0.01$), supporting our hypothesis of woodburning as the primary source of model underprediction in this region. No significant spatial autocorrelation existed in LUR model residuals for NO_2 (Moran's $I = 0.03$, $p = 0.44$), benzene (Moran's $I = -0.20$, $p = 0.13$), ethylbenzene (Moran's $I = -0.00$, $p = 0.87$), and 1,3-butadiene (Moran's $I = 0.09$, $p = 0.32$).

Incorporating gradients with national LUR models. Deterministic gradients were added to LUR models, because we could not estimate the effects of local-scale pollution sources from NAPS data alone. Figure 2A illustrates the final $PM_{2.5}$ model (LUR plus gradients) for Canada as a whole and for southern Ontario and the city of Toronto. Figure 2B illustrates the final national NO_2 model (LUR plus gradients) for Canada as a whole and for southwestern British Columbia and the city of Vancouver. These maps illustrate the spatial resolution of the final national pollutant models; however, for population exposure assessment, the LUR model results and deterministic gradients were applied to street block point locations, as shown in Figure 3, which illustrates the final national benzene model (LUR plus gradients) calculated at the block point level.

Evaluation of national pollutant models. Evaluation of LUR models using bootstrap and leave-one-out analyses. The distribution of all model coefficients for each pollutant resulting from bootstrap analysis showed normal distributions. The NO_2 model was the least sensitive to monitor selection, with a bootstrap R^2 95% CI of 65–81. Models for $PM_{2.5}$, benzene, ethylbenzene, and 1,3-butadiene demonstrated larger uncertainty to monitor selection, with R^2 95% CIs of 33–59, 44–80, 49–85, and 53–82, respectively. Variable coefficients for industrial NPRI proximity variables were extremely sensitive to monitor selection. The leave-one-out analyses indicated no significant bias in any LUR model, as demonstrated by the mean \pm SD error: $1.07e-3 \pm 5.61$ for NO_2 ; $-6.35e-3 \pm 1.59$ for $PM_{2.5}$; -0.04 ± 0.32 for benzene; -0.01 ± 0.04 for 1,3-butadiene; and -0.04 ± 0.22 for ethylbenzene.

Evaluation of NO_2 and benzene models using city-specific data. On average, the national NO_2 LUR plus gradient model predicted 43% of the within-city NO_2 variation (based on the city-specific data evaluation)

compared with 22% predicted based on IDW of NAPS monitors plus gradients (Table 3). National LUR, LUR plus gradients, IDW, and IDW plus gradients models overpredicted the city-specific NO_2 measurements, with average city-specific intercepts of 4.56, 7.45, 8.51, and $11.56 \mu\text{g}/\text{m}^3$, respectively. City-specific scatter plots of measured and modeled NO_2 concentrations are illustrated in Supplemental Material, Figure 1 (doi:10.1289/ehp.1002976).

For benzene, all modeling methods performed poorly in explaining within-city benzene variation. The LUR plus gradients model explained, on average, only 16% of within-city variability in benzene concentrations compared with 11% based on IDW plus gradients (Table 3). In the evaluation using the Montreal city-specific benzene concentrations, four outliers were removed (all concentrations $> 2 \mu\text{g}/\text{m}^3$), and one outlier ($4.10 \mu\text{g}/\text{m}^3$) was removed in the Toronto evaluation. Benzene models also overpredicted city-specific concentrations, based on city-specific intercepts of modeled versus measured concentrations [see Supplemental Material Figure 2 (doi:10.1289/ehp.1002976)]. Sarnia, a high-density

industrial community with 46 NPRI emitters, had poor NO_2 and benzene model evaluations.

Canadian population exposure assessment. The final LUR models and gradients were applied to all 478,831 street block-face centroid locations to conduct population exposure assessments. Estimated mean (95% CI) population exposures (micrograms per cubic meter) to ambient $PM_{2.5}$, NO_2 , benzene, ethylbenzene, and 1,3-butadiene in Canada based on the LUR models were 8.10 (5.84–10.43), 22.40 (13.14–33.51), 0.94 (0.57–1.31), 0.38 (0.25–0.52), and 0.086 (0.035–0.138), respectively. Estimates for the same pollutants based on the national LUR plus gradients models were 8.39 (6.00–11.13), 23.37 (14.01–35.73), 1.04 (0.59–1.49), 0.63 (0.35–1.10), and 0.089 (0.036–0.146), respectively. Wide ranges of exposure levels were estimated in Canada for all substances; see Supplemental Material, Figure 3 (doi:10.1289/ehp.1002976) for population exposure distributions.

Discussion

We created national pollutant models from fixed-site monitoring data that incorporate

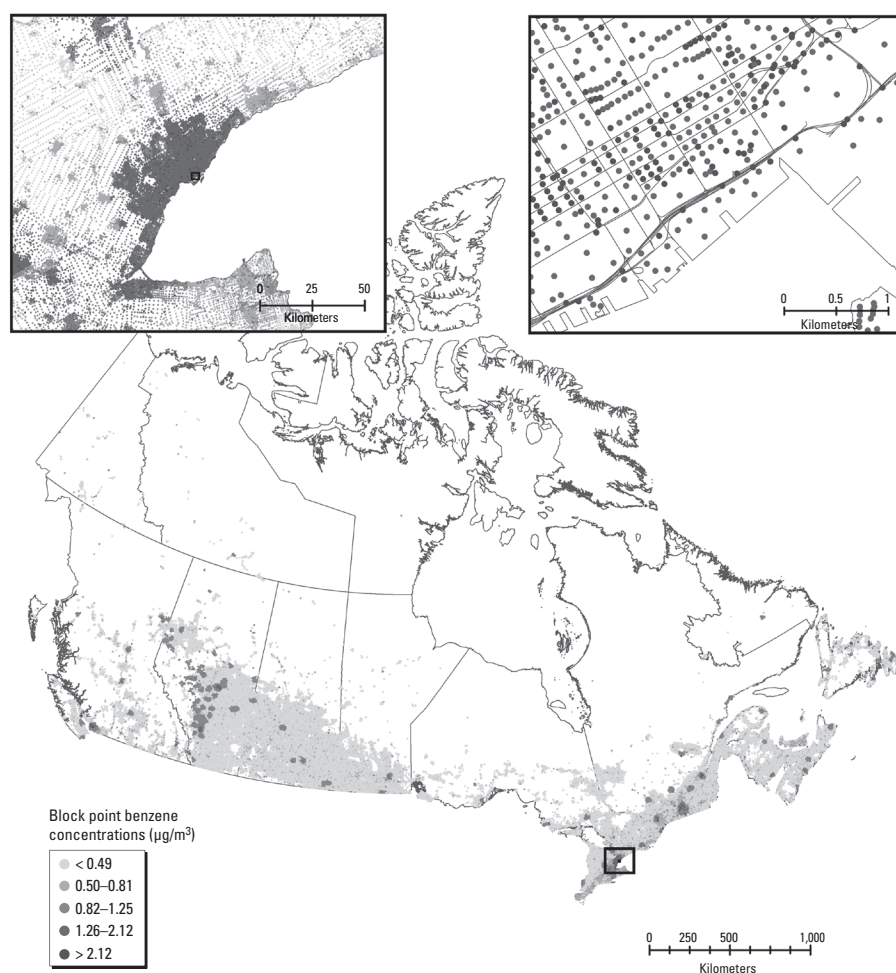


Figure 3. National benzene LUR model plus gradients (illustrating the city of Toronto) calculated for each street block point in Canada ($n = 478,831$).

satellite, geographic, and deterministic components and demonstrated that these models can improve exposure assessment over large geographic areas compared with approaches based solely on interpolation of fixed-site monitoring data. We also demonstrated how these models can be used for population exposure assessment.

The national LUR models explained 73% of pollution variation in NAPS measurements for NO₂, and lesser degrees for PM_{2.5} (46%), benzene (62%), ethylbenzene (67%), and 1,3-butadiene (68%). The NO₂ and PM_{2.5} models were least sensitive to monitor selection, whereas models for VOCs were more sensitive—likely because of the smaller number of monitors on which LUR estimates were based ($n = 53$). The predictive performance of the PM_{2.5} model [$R^2 = 0.46$, root mean square error (RMSE) = 1.53 µg/m³] was consistent with other large-scale modeling studies based on different monitoring methodologies and data inputs (Beelen et al. 2009; Hart et al. 2009; Liao et al. 2006; Ross et al. 2007).

The national LUR models generally captured regional patterns in pollutant concentrations, corresponding to NAPS monitor siting criteria, but were less effective at identifying small-scale geographic predictor variables. For example, only 35 NAPS monitors were located within 500 m of a major road and only 7 monitors were within 500 m of a major industrial emission source. Such small sample sizes greatly reduce the power of the models to capture these specific pollutant sources. Some city-specific LUR methods have used location-allocation methods to more fully represent the true spatial variation in pollution levels and to capture the range of predictor variables (Jerrett et al. 2005). Models based on fixed-site monitor data may therefore need additional approaches to represent local-scale pollutant

variability not captured by fixed-site monitors. This was indeed the case with the Canadian NAPS network, but larger regulatory networks, such as those in the United States, may better represent the range of predictor variables needed to build local-scale LUR models.

To address the lack of local-scale geographic variability in the NAPS data, we incorporated deterministic gradients based on proximity to specific sources (i.e., vehicles and gas stations). The final NO₂ LUR plus gradient model improved prediction of within-city pollutant variation considerably compared with the LUR model alone and interpolation methods. On average, the final model predicted 43% of within-city NO₂ variation compared with 18% using IDW. Both the national benzene model and IDW predicted within-city benzene poorly, which may be due to the small number of NAPS monitors on which the model was based, the relatively small variation in within-city benzene levels, or the inability of gradients to capture local benzene concentrations. Similar to the NO₂ model, the evaluation of the benzene model with Sarnia data was poor, reflecting the difficulty in capturing unique high-density industrial conditions in a national model.

Gradients were based on literature reviews. The lack of methodological consistency among published data of pollutant level increases near specific sources and the distance required for pollutant levels to return to background were clear limitations. To improve reliability of gradients, we used linear functions to represent the decreases in pollutant levels found in the initial portions of the exponential decay curves found in the literature. The methodology used here could be augmented as new gradients become available or with other modeled data.

Population exposure assessment was conducted using the national models and

census street block-face points. The population-weighted average exposures to PM_{2.5}, NO₂, benzene, ethylbenzene, and 1,3-butadiene were 8.39, 23.37, 1.04, 0.63, and 0.089 (µg/m³), respectively. The uncertainty of population exposure estimates were driven primarily by LUR model uncertainty. Although the results of the national LUR models are similar to city-specific LUR models in their predictive capacity and error, we are unaware of any LUR models that have been applied to estimate exposure uncertainty. Although these exposures are low compared with other developed countries, exposures in particular locations in Canada are relatively high. For example, the 90th percentiles of exposures (micrograms per cubic meter) are 9.78 for PM_{2.5}, 34.81 for NO₂, 1.61 for benzene, 1.01 for ethylbenzene, and 0.14 for 1,3-butadiene. The ability of the national models to capture local-scale pollutant variability allows for more realistic exposure assessments and assessments that can potentially identify high-risk populations. Future work will refine approaches for using the national models to calculate population exposure assessments, incorporate socioeconomic information from census to examine environment injustice issues, and integrate national models into a risk-assessment framework that incorporates exposures from other sources and microenvironments.

This study faced a number of challenges and limitations to creating national pollutant models from fixed-site monitors and applying these models to estimate Canadian population exposures. First, the NAPS monitors in Canada are centered in large metropolitan areas, and modeled relationships will therefore be weighted toward these areas. This is appropriate for population exposure assessment, because these locations represent the majority of Canadians, but in rural areas the models could be adjusted or a background concentration could be used. This is particularly relevant to the benzene, ethylbenzene, and 1,3-butadiene models, which were based on data from monitors located almost exclusively in large urban areas or sited near large industrial sources. Second, we had limited data on pollutant sources and source strengths such as traffic volumes. In addition, we did not model emissions from woodburning stoves and forest fires, which may have caused us to underpredict PM_{2.5} concentrations in the interior of British Columbia. Third, parsimonious LUR models were created, because the specificity of model variables may be important for informing surveillance and regulation. This, however, leads to models that do not capture the complex interactions between geographic characteristics and pollutant sources, and even the simplest LUR predictors (e.g., major roads or NPRI facilities within 10 km) capture complex mixes of geographic characteristics and pollutant sources. Fourth, we compared model estimates with city-specific measurements

Table 3. Evaluation of national NO₂ and benzene models, as well as IDW estimates from fixed-site monitors, against independent city-specific measurement data.

| Substance | n^a | R^2 (RMSE) | | | |
|-----------------------|-------|------------------|----------------------|------------------|----------------------|
| | | LUR ^b | LUR + G ^c | IDW ^d | IDW + G ^e |
| NO ₂ | | | | | |
| Edmonton | 50 | 0.60 (3.67) | 0.41 (4.59) | 0.10 (5.52) | 0.01 (5.92) |
| Montreal | 135 | 0.41 (4.28) | 0.48 (4.04) | 0.31 (4.63) | 0.41 (4.29) |
| Sarnia | 34 | 0.42 (4.21) | 0.49 (4.04) | 0.12 (5.15) | 0.19 (5.12) |
| Toronto | 196 | 0.18 (7.69) | 0.36 (6.78) | 0.13 (7.93) | 0.32 (6.99) |
| Victoria | 40 | 0.19 (3.95) | 0.37 (3.70) | 0.23 (3.86) | 0.26 (3.98) |
| Vancouver | 114 | 0.31 (6.41) | 0.42 (5.93) | 0.31 (6.43) | 0.36 (6.24) |
| Winnipeg | 49 | 0.54 (3.65) | 0.51 (3.86) | 0.08 (5.17) | 0.02 (5.43) |
| Average | 618 | 0.39 (4.84) | 0.43 (4.71) | 0.18 (5.53) | 0.22 (5.42) |
| Benzene | | | | | |
| Montreal ^f | 131 | 0.33 (0.24) | 0.26 (0.25) | 0.11 (0.28) | 0.05 (0.29) |
| Sarnia | 37 | 0.02 (0.57) | 0.04 (0.56) | 0.00 (0.57) | 0.03 (0.56) |
| Toronto ^g | 44 | 0.03 (0.19) | 0.22 (0.17) | 0.00 (0.19) | 0.34 (0.16) |
| Winnipeg | 94 | 0.08 (0.25) | 0.10 (0.25) | 0.00 (0.26) | 0.01 (0.26) |
| Average | 306 | 0.12 (0.31) | 0.16 (0.31) | 0.03 (0.33) | 0.11 (0.32) |

^aNumber of within-city measurement locations. ^bNational LUR model. ^cNational LUR model plus gradients (G). ^dIDW interpolation of NAPS fixed-site monitoring data. ^eIDW interpolation of NAPS fixed-site monitoring data plus gradients. ^fFour outliers removed with highest city concentrations (> 2 µg/m³). ^gOne outlier removed with highest city concentration (4.10 µg/m³).

for NO₂ and benzene collected in different years and using a variety of methodologies. Nevertheless, these measurements represent the best data on within-city pollutant variability available. Fifth, applying LUR model results to approximately half a million block points is currently extremely computationally and time intensive. Finally, the geographic accuracy of street block centroids may introduce errors into the gradient portions of the models and therefore the exposure assessment, particularly between rural and urban areas. These errors, however, are likely spatially random within rural and urban areas across Canada.

Conclusions

National exposure models were required by Carex Canada to produce population exposure assessments that captured both between-city and within-city pollution variability. We created national PM_{2.5}, NO₂, benzene, ethylbenzene, and 1,3-butadiene models from fixed-site monitoring data and found that a combination of data sources and methods to capture background, regional, and local-scale pollution variation improved exposure assessment over traditional IDW interpolation approaches. The national pollutant models were applied to street block-face points, representing the locations of the Canadian population, to determine population exposure estimates. Estimates of average population exposure levels in Canada are PM_{2.5} 8.39, NO₂ 23.37, benzene 1.04, ethylbenzene 0.63, and 1,3-butadiene 0.09 (µg/m³). The modeling approach developed here uses readily available data and could be reproduced over time, for example, every 5 years with the Canadian census. This would provide updated population exposure assessments and a long-term surveillance capacity for monitoring trends in population exposures, for identifying potential susceptible populations and geographic locations with elevated exposures, and for evaluating the impacts of policies and regulatory changes on exposure levels.

REFERENCES

- Allen RW, Amram O, Wheeler A, Brauer M. 2010. The transferability of NO and NO₂ land use regression models between cities and pollutants. *Atmos Environ* 45(2):369–378.
- Atari DO, Luginaah IN. 2009. Assessing the distribution of volatile organic compounds using land use regression in Sarnia, "Chemical Valley," Ontario, Canada. *Environ Health* 8:16; doi:10.1186/1476-069X-8-16 [Online 16 April 2009].
- Beckerman B, Jerrett M, Brook JR, Verma DK, Arain MA, Finkelstein MM. 2008. Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. *Atmos Environ* 42(2):275–290.
- Beelen R, Hoek G, Pebesma E, Vienneau D, de Hoogh K, Briggs DJ. 2009. Mapping of background air pollution at a fine spatial scale across the European Union. *Sci Total Environ* 407(6):1852–1867.
- Carex Canada 2011. Surveillance of Environmental and Occupational Exposures for Cancer Prevention. Carcinogen Database. Available: http://www.carexcanada.ca/en/our_research [accessed 15 January 2011].
- Crouse DL, Goldberg MS, Ross NA. 2009. A prediction-based approach to modeling temporal and spatial variability of traffic-related air pollution in Montreal, Canada. *Atmos Environ* 43(32):5075–5084.
- Cyrys J, Hochadel M, Gehring U, Hoek G, Diegmann V, Brunekreef B, et al. 2005. GIS-based estimation of exposure to particulate matter and NO₂ in an urban area: stochastic versus dispersion modeling. *Environ Health Perspect* 113:987.
- Environment Canada. 2010. National Pollutant Release Inventory. Tracking Pollution in Canada. Available: <http://www.ec.gc.ca/inrp-npri/default.asp?lang=en> [accessed 17 January 2011].
- GEOS-Chem. 2011. GEOS-CHEM Model. Available: <http://acmg.seas.harvard.edu/geos/> [accessed 15 January 2011].
- Gilbert NL, Goldberg MS, Brook JR, Jerrett M. 2007. The influence of highway traffic on ambient nitrogen dioxide concentrations beyond the immediate vicinity of highways. *Atmos Environ* 41(12):2670–2673.
- Gilbert NL, Woodhouse S, Stieb DM, Brook JR. 2003. Ambient nitrogen dioxide and distance from a major highway. *Sci Total Environ* 312(1–3):43–46.
- Global Land Cover Characterization. 2008. Global Land Cover Characterization Background. Available: <http://edc2.usgs.gov/glcc/background.php> [accessed 3 October 2010].
- Google Scholar. 2010. Advanced Scholar Search Homepage. Available: http://scholar.google.com/advanced_scholar_search [accessed 24 November 2010].
- Hart JE, Yanosky JD, Puett RC, Ryan L, Dockery DW, Smith TJ, et al. 2009. Spatial modeling of PM₁₀ and NO₂ in the continental United States, 1985–2000. *Environ Health Perspect* 117:1690–1696.
- Hellén H, Hakola H, Pirjola L, Laurila T, Pystynen KH. 2006. Ambient air concentrations, source profiles, and source apportionment of 71 different C2–C10 volatile organic compounds in urban and residential areas of Finland. *Environ Sci Technol* 40(1):103–108.
- Henderson SB, Beckerman B, Jerrett M, Brauer M. 2007. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ Sci Technol* 41(7):2422–2428.
- Hitchins J, Morawska L, Wolff R, Gilbert D. 2000. Concentrations of submicrometre particles from vehicle emissions near a major road. *Atmos Environ* 34(1):51–59.
- Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, et al. 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ* 42(33):7561–7578.
- Jerrett M, Arain M, Kanaroglou P, Beckerman B, Crouse D, Gilbert N, et al. 2007. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *J Toxicol Environ Health A* 70(3–4):200–212.
- Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahuvaroglu T, et al. 2005. A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol* 15(2):185–204.
- Karakitsios SP, Delis VK, Kassomenos PA, Pilidis GA. 2007. Contribution to ambient benzene concentrations in the vicinity of petrol stations: estimation of the associated health risk. *Atmos Environ* 41(9):1889–1902.
- Lamsal L, Martin R, van Donkelaar A, Steinbacher M, Celarier E, Bucsela E, et al. 2008. Ground-level nitrogen dioxide concentrations inferred from the satellite-borne ozone monitoring instrument. *J Geophys Res* 113:D16308; doi:10.1029/2007JD009235 [Online 28 August 2008].
- Liao D, Pequet DJ, Duan Y, Whitsel EA, Dou J, Smith RL, et al. 2006. GIS approaches for the estimation of residential-level ambient PM concentrations. *Environ Health Perspect* 114:1374–1380.
- Liu Y, Paciorek CJ, Koutrakis P. 2009. Estimating regional spatial and temporal variability of PM_{2.5} concentrations using satellite data, meteorology, and land use information. *Environ Health Perspect* 117:886–892.
- Martin RV. 2008. Satellite remote sensing of surface air quality. *Atmos Environ* 42(34):7823–7843.
- Mukerjee S, Smith LA, Johnson MM, Neas LM, Stallings CA. 2009. Spatial analysis and land use regression of VOCs and NO₂ from school-based urban air monitoring in Detroit/Dearborn, USA. *Sci Total Environ* 407(16):4642–4651.
- Nafstad P, Haheim LL, Oftedal B, Gram F, Holme I, Hjermann I, et al. 2003. Lung cancer and air pollution: a 27 year follow up of 16,209 Norwegian men. *Thorax* 58(12):1071–1076.
- NAPS (National Air Pollution Surveillance Network). 2004. NAPS Quality Assurance and Quality Control Guidelines. Report No. AAQD 2004-1. Available: <http://www.etc-cte.ec.gc.ca/publications/naps/NAPSQAQC.pdf> [accessed 15 January 2011].
- NASA (National Aeronautics and Space Administration). 2011a. Aura Satellite. Available: <http://aura.gsfc.nasa.gov/> [accessed 23 May 2011].
- NASA (National Aeronautics and Space Administration). 2011b. Terra Satellite. Available: <http://terra.nasa.gov/> [accessed 3 May 2011].
- Parra M, Elustondo D, Bermejo R, Santamaria J. 2009. Ambient air levels of volatile organic compounds (VOC) and nitrogen dioxide (NO₂) in a medium size city in northern Spain. *Sci Total Environ* 407(3):999–1009.
- PubMed Central. 2010. PubMed Central Homepage. Available: <http://www.pubmedcentral.nih.gov/> [accessed 15 November 2010].
- Roorda-Knappe MC, Janssen NAH, De Hartog JJ, Van Vliet PHN, Harssema H, Brunekreef B. 1998. Air pollution from traffic in city districts near major motorways. *Atmos Environ* 32(11):1921–1930.
- Ross Z, Jerrett M, Ito K, Tempalski B, Thurston GD. 2007. A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmos Environ* 41(11):2255–2269.
- Roukos J, Riffault V, Locoge N, Plaisance H. 2009. VOC in an urban and industrial harbor on the French North Sea coast during two contrasted meteorological situations. *Environ Pollut* 157(11):3001–3009.
- Smargiassi A, Baldwin M, Pilger C, Dugandzic R, Brauer M. 2005. Small-scale spatial variability of particle concentrations and traffic levels in Montreal: a pilot study. *Sci Total Environ* 338(3):243–251.
- Smith L, Mukerjee S, Gonzales M, Stallings C, Neas L, Norris G, et al. 2006. Use of GIS and ancillary variables to predict volatile organic compound and nitrogen dioxide levels at unmonitored locations. *Atmos Environ* 40(20):3773–3787.
- Statistics Canada. (2006). Block-face representative points. Available: <http://www12.statcan.ca/census-recensement/2006/ref/dict/geo040a-eng.cfm> [accessed 3 October 2011].
- Su J, Jerrett M, Beckerman B, Verma D, Altaf Arain M, Kanaroglou P, et al. 2010. A land use regression model for predicting ambient volatile organic compound concentrations in Toronto, Canada. *Atmos Environ* 44(29):3529–3537.
- Su JG, Jerrett M, Beckerman B, Wilhelm M, Ghosh JK, Ritz B. 2009. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. *Environ Res* 109(6):657–670.
- Thomson Reuters. 2010. ISI Web of Knowledge Central Homepage. Available: http://apps.isiknowledge.com/WOS_GeneralSearch_input.do?highlighted_tab=WOS&product=WOS&lastprod=WOS&search_mode=GeneralSearch&SID=1AjaopJLKe4Mb31h29l [accessed 9 November 2010].
- Thorsson S, Eliasson I. 2006. Passive and active sampling of benzene in different urban environments in Gothenburg, Sweden. *Water Air Soil Pollut* 173(1):39–56.
- Tiitta P, Raunemaa T, Tissari J, Yli-Tuomi T, Leskinen A, Kukkonen J, et al. 2002. Measurements and modeling of PM_{2.5} concentrations near a major road in Kuopio, Finland. *Atmos Environ* 36(25):4057–4068.
- van Donkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C, et al. 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environ Health Perspect* 118:847–855.
- Vardoulakis S, Gonzalez-Flesca N, Fisher B. 2002. Assessment of traffic-related air pollution in two street canyons in Paris: implications for exposure studies. *Atmos Environ* 36(6):1025–1039.
- Venkatram A, Isakov V, Seila R, Baldauf R. 2009. Modeling the impacts of traffic emissions on air toxics concentrations near roadways. *Atmos Environ* 43(20):3191–3199.
- Wang P, Zhao W. 2008. Assessment of ambient volatile organic compounds (VOCs) near major roads in urban Nanjing, China. *Atmos Res* 89(3):289–297.
- Wheeler AJ, Smith-Doiron M, Xu X, Gilbert NL, Brook JR. 2008. Intra-urban variability of air pollution in Windsor, Ontario: measurement and modeling for human exposure assessment. *Environ Res* 106(1):7–16.
- Yanosky JD, Paciorek CJ, Schwartz J, Laden F, Puett R, Suh HH. 2008. Spatio-temporal modeling of chronic PM₁₀ exposure for the Nurses' Health Study. *Atmos Environ* 42(18):4047–4062.